Universidade de Brasília
Instituto de Ciências Exatas
Departamento de Estatística

COMPARANDO O ESTIMADOR DO TIPO RAZÃO MULTIVARIADO COM O ESTIMADOR DO TIPO REGRESSÃO

Andre Ricardo Santana da Costa **10/0048218**

Brasília

2013

Andre Ricardo Santana da Costa **10/0048218**

COMPARANDO O ESTIMADOR DO TIPO RAZÃO MULTIVARIADO COM O ESTIMADOR DO TIPO REGRESSÃO

Relatório apresentado à disciplina Estágio Supervisionado II do curso de graduação em Estatística, Departamento de Estatística, Instituto de Exatas, Universidade de Brasília, como parte dos requisitos necessários para o grau de Bacharel em Estatística.

Orientador: Prof. Dr. Alan Ricardo da Silva

Brasília

2013

Dedico este trabalho à Deus, à minha família e em especial à minha esposa.

Agradecimentos

Em primeiro lugar agradeço a minha família por me apoiar e em especial à minha esposa por estar ao meu lado nos momentos de dificuldade e de angústias, me ajudando sempre a tomar as melhores decisões de minha vida. Agradeço também ao professor Alan Ricardo da Silva por ter paciência e disponibilidade para tirar dúvidas e me ajudar a desenvolver meu primeiro trabalho estatístico.

"A mente que se abre a uma nova idéia jamais voltará ao seu tamanho original." Albert Einstein

Resumo

Muita das vezes não é possível estimar uma variável de interesse diretamente, somente através de 1 ou mais varíaveis auxiliares. Um exemplo clássico é a estimação da quantidade de água que pode ser extraída de um caminhão carregado de cocos. As unidades populacionais são os cocos e a variável de interesse é o total de água no carregamento. O estimador natural seria o estimador expansão ($\hat{\tau}_Y = T_Y = N\overline{y}$), porém não possível sua utilização porque o número de cocos no carregamento não é conhecido. Contudo sabe-se que o peso é fortemente correlacionado com a quantidade da água no coco. Então pode-se definir a razão, quantidade média de água por unidade de peso. Sabendo disso, a quantidade de água no carregamento será igual a razão multiplicado pelo peso total do carregamento. Tem-se então o Estimador do Tipo Razão Univariado, ou seja, através de uma variável auxiliar (peso do total carregamento) pode-se estimar a variável de interesse (quantidade de água no carregamento).

Contudo, algumas vezes pode-se ter mais de uma variável auxiliar para a determinação da variável de interesse, neste caso recomenda-se a utilização do Estimador do Tipo Razão Multivariado ou ainda o Estimador do Tipo Regressão Multivariado que tem como um caso particular o Estimador do Tipo Razão. Portanto o intuito do estudo é comparar a eficiência dos 2 estimadores e saber qual é mais adequado em determinadas situações, e como o software SAS pode auxiliar na obtenção das estimativas de interesse. No estudo será possível verificar que o Estimador do Tipo Razão Multivariado pode ser mais adequado quando se conhece todas as variáveis auxiliares X_i's para estimar a variável de interesse Y. Se não houver o conhecimento de todas, na maioria dos casos o Estimador do Tipo Regressão Multivariado será

mais adequado.

Palavras-chaves: amostragem, estimador, razão, regressão, estimador do tipo razão, estimador do tipo regressão, estimador do tipo razão multivariado

Lista de Tabelas

Lista de Figuras

Sumário

Capítulo 1

INTRODUÇÃO

A escolha do melhor plano amostral depende diretamente da estrutura como os indivíduos estão distribuídos na população e como podem ser coletados. O Estimador do Tipo Razão é apropriado quando a variável de interesse só pode ser estimada através de uma variável auxiliar ou preditora, ou para melhorar a previsão de parâmetros como a média ou o total populacional, portanto tem-se um par $(X_i, Y_i), i = 1, ..., N$, associado a cada elemento da população. Há casos em que uma variável de interesse possui p variáveis auxiliares, neste caso estamos lidando com o Estimador do Tipo Razão Multivariado que é descrito da seguinte forma:

$$\begin{aligned} \hat{Y}_{RM} &= W_1 \frac{\overline{y}}{\overline{x}_1} X_1 + W_2 \frac{\overline{y}}{\overline{x}_2} X_2 + W_3 \frac{\overline{y}}{\overline{x}_3} X_3 + ... + W_p \frac{\overline{y}}{\overline{x}_p} X_p \\ &= W_1 \hat{Y}_{R_1} + W_2 \hat{Y}_{R_2} + W_3 \hat{Y}_{R_3} + ... + W_p \hat{Y}_{R_p} \end{aligned} \tag{1.1}$$

onde W_j é o peso da j-ésima variável auxiliar, e a soma $\sum_j^p W_j = 1$ maximiza a precisão de $\hat{Y}_{RM}$. O Estimador do Tipo Razão é um caso particular do Estimador do Tipo Regressão quando a reta de regressão da relação entre Y e X_j passa pela origem. Dessa forma, esse trabalho visa apresentar a estrutura do Estimador Tipo Razão Multivariado além de comparar sua eficiência com o Estimador do Tipo Regressão.

1.1 OBJETIVOS

O objetivo geral do trabalho é implementar o Estimador do Tipo Razão Multivariado no *software* SAS.
Os objetivos específicos são:

- conduzir sua implementação via PROC IML no *software* SAS 9.2 para os principais planos amostrais;
- mostrar o poder e a aplicabilidade do Estimador do Tipo Razão Multivariado;
- comparar a eficiência do Estimador do Tipo Razão Multivariado com o Estimador do Tipo Regressão Multivariado.

Capítulo 2

ESTIMADOR DO TIPO RAZÃO

2.1 INTRODUÇÃO

A amostragem é uma das áreas da estatísticas mais utilizadas, pois para que se possa realizar um trabalho de qualidade é preciso que os dados sejam coletados com cuidado para que as estimativas obtidas tenham boa precisão. Em alguns casos não é possível estimar o total populacional de uma variável diretamente, então é necessário utilizar uma variável auxiliar para a estimação. Neste caso é possível utilizar o Estimador do Tipo Regressão ou o Estimador do Tipo Razão. Este Capítulo fará uma breve revisão no Estimador do Tipo Razão univariado e em seguida apresentará o Estimador do Tipo Razão Multivariado que é foco do estudo.

2.2 ESTIMADOR DO TIPO RAZÃO UNIVARIADO

No caso univariado a variável de interesse Y é não conhecida, porém ela é estimada por uma variável auxiliar X conhecida. Então tem-se uma amostra aleatória $(x_1, y_1)...(x_n, y_n)$ de uma população $(X_1, Y_1)...(X_N, Y_N)$. De acordo com Bolfarine and Bussab (2005), a média $\overline{X}$ é conhecida e através delas é possível estimar $\overline{Y}$ da seguinte forma:

$$\hat{\overline{Y}} = \frac{\overline{y}}{\overline{x}}\overline{X} = r\overline{X} \tag{2.1}$$

que é o Estimador do Tipo Razão de $\overline{Y}$. Sabe-se que $\hat{\overline{Y}}$ é viesado para n pequeno, porém para amostras grandes a estimativa é boa. Neste caso tem-se:

$$E(\hat{\overline{Y}}) = \overline{Y} + \frac{(N-n)}{N}\frac{\overline{Y}}{n}(c_{xx} - cxy), \tag{2.2}$$

$$Var(\hat{\overline{Y}}) = \frac{(N-n)}{N}\frac{\overline{Y}^2}{n}(c_{xx} + c_{yy} - 2c_{xy}), \tag{2.3}$$

onde $c_{xx} = \dfrac{S_{xx}}{\overline{X}^2}$, $c_{yy} = \dfrac{S_{yy}}{\overline{Y}^2}$ e $c_{xy} = \dfrac{S_{xy}}{\overline{XY}}$, sendo que S_{xx} e S_{yy} correspondem à variância amostral e S_{xy} a covariância entre X e Y. Hartley and Ross (1954) demonstraram que o estimador não-viesado de $\hat{\overline{Y}}$ é dado por:

$$\hat{\overline{Y}}^* = \overline{r}\overline{X} + \frac{(N-1)n(\overline{y} - \overline{rx})}{(n-1)N} \tag{2.4}$$

no qual $n\overline{r} = \Sigma y_i/x_i$, para $x_i > 0$. De acordo com Cochran (1977) $\hat{\overline{Y}} \to \overline{Y}$ quando $n \to N$, ou seja, $\hat{\overline{Y}}$ é um estimador consistente de $\overline{Y}$.

2.3 ESTIMADOR DO TIPO RAZÃO MULTIVARIADO

Extendendo o Estimador do Tipo Razão para o caso Multivariado tem-se o seguinte modelo para a população:

$$\begin{gathered} Y_1,...,Y_N,\ \overline{Y} \text{ não conhecido} \\ X_11,...,X_1N,\ \overline{X}_1 \neq 0 \text{ e conhecido, } R_1 = \frac{\overline{Y}}{\overline{X}_1}, \\ .... \\ X_p1,...,X_1N,\ \overline{X}_p \neq 0 \text{ e conhecido, } R_p = \frac{\overline{Y}}{\overline{X}_p}, \end{gathered}$$

De acordo com Olkin (1958) a matrix de covariância $(p+1)\times(p+1)$ também é conhecida e os índices 0, 1,..., p referem-se às variáveis Y, X_1,..., X_p respectivamente. Os momentos são dados por:

$$\mu_{12}^{ij} = \frac{\sum_k (X_{ik} - \overline{X}_i)(X_{jk} - \overline{X}_j)^2}{N} \tag{2.5}$$

$$\mu_{111}^{0ij} = \frac{\sum_k (Y_k - \overline{Y})(X_{ik} - \overline{X}_i)(X_{jk} - \overline{X}_j)}{N} \tag{2.6}$$

onde os índices superiores indicam as variáveis e os inferiores o poder. Então a covariância é dada por $S_{ij} = N\mu_{11}^{ij}/(N-1)$ e o coeficiente de variação dado por

$c_i = S_i/\overline{X}_i$. Dividindo os momentos pelas médias tem-se o peso das variáveis X_i´s:

$$w_{12}^{ij} = \frac{\mu_{12}^{ij}}{\overline{X}_i \overline{X}_i^2} \tag{2.7}$$

Ao se retirar uma amostra aletória simples $(y_j, x_{1j}, ..., x_{pj})$ de uma população onde $j = 1, 2, ..., n$, é obtido o Estimador do Tipo Razão Multivariado de $\overline{Y}$ dado por:

$$\hat{\overline{Y}} = w_1 r_1 \overline{X}_1 + ... + w_p r_p \overline{X}_p \tag{2.8}$$

onde $\mathbf{w} = (w_1, w_2, ..., w_p)$ são os pesos sob a restrição $\Sigma w_i = 1$ e $r_i = \overline{y}/\overline{x}_i$. Assim como no caso univariado, este estimador é viesado para amostras pequenas. Então um outro Estimador dado por Hartley and Ross (1954) é:

$$\hat{\overline{Y}}^* = \sum_1^p w_i r_i \overline{X}_i + \frac{(N-1)n}{N(n-1)}(\overline{y} - \sum_1^p w_i \overline{r}_i \overline{x}_i) \tag{2.9}$$

que é um estimador não viesado de $\overline{Y}$, onde $n\overline{r}_i = \sum_1^n y_j/x_{ij}$. Tem-se aqui uma combinação linear de estimadores consistentes. A média e a variância de $\hat{y}$ é dado por:

$$E\hat{\overline{Y}} = \overline{Y} \Sigma w_i E(r_i/R_i), \tag{2.10}$$

$$Var(\hat{\overline{Y}}) = \overline{Y}^2 \Sigma w_i w_j cov(r_i, r_j)/R_i R_j. \tag{2.11}$$

Para estimar $E\hat{\overline{Y}}$ e $Var(\hat{\overline{Y}})$, Olkin (1958) utilizou o método delta e obteve o seguinte resultado:

$$E\hat{\overline{Y}} = \overline{Y} + \overline{Y}\mathbf{w}\mathbf{b}', \tag{2.12}$$

$$Var(\hat{\overline{Y}}) = \frac{\overline{Y}^2}{n}\mathbf{w}\mathbf{A}\mathbf{w}' \tag{2.13}$$

onde os vetores $\mathbf{b} = (b_1, ..., b_p)$ e $\mathbf{a} = (a_1, ..., a_p)$ e a matriz $\mathbf{A} = (a_{ij}) : p \times p$ definidos por:

$$b_i = \frac{N-n}{N}(c_i^2 - \rho_{0i} c_0 c_i) \tag{2.14}$$

$$a_{ij} = \frac{N-n}{N}(c_0^2 - \rho_{0i} c_0 c_i - \rho_{0j} c_0 c_j + \rho_{ij} c_i c_j) \tag{2.15}$$

com ρ_{ij} sendo a correlação e c_i o coeficiente de variação. Esses vetores, de acordo com Olkin (1958), podem ser estimados da seguinte forma:

$$\hat{b}_i = (\frac{N-n}{N})\frac{\sum_{t=1}^n y_t(y_t - r_i x_{it})}{n-1} \tag{2.16}$$

$$\hat{a}_{ij} = (\frac{N-n}{N})\frac{\sum_{t=1}^n (y_t - r_i x_{it})(y_t - r_j x_{jt})}{n-1} \tag{2.17}$$

Para escolher os pesos das variáveis auxiliares deve-se levar em conta alguns pontos. Primeiramente $\Sigma w_i = 1$, segundo escolher os w_i's de forma que minimize a variância de $\hat{\overline{Y}}$. Para estimar os pesos é necessário utilizar a desigualdade de Cauchy, gerando o seguinte resultado:

$$\hat{\mathbf{w}} = \frac{\mathbf{eA}^{-1}}{\mathbf{eA}^{-1}\mathbf{e}'} \tag{2.18}$$

onde $\mathbf{e}$ é o vetor $\{1, 1, ..., 1\}$. Então o valor esperaro e a variância de $\hat{\overline{Y}}$ será:

$$E\hat{\overline{Y}} = \overline{Y} + \frac{\overline{Y}}{n}\frac{\mathbf{eA}^{-1}\mathbf{b}'}{\mathbf{eA}^{-1}\mathbf{e}'} \tag{2.19}$$

$$Var(\hat{\overline{Y}}) = \frac{\overline{Y}^2}{n}\frac{1}{\mathbf{eA}^{-1}\mathbf{e}'} \tag{2.20}$$

De acordo com (Olkin, 1958), esses parâmetros podem ser estimados da seguinte forma:

$$E\hat{\overline{y}} = \sum_{i=1}^{p} \hat{w}_i \hat{R}_i \overline{x}_i \tag{2.21}$$

$$Var(\hat{\overline{y}}) = \frac{1}{n(\mathbf{e}\hat{\mathbf{A}}^{-1}\mathbf{e}')} \tag{2.22}$$

no qual os elementos da matriz $\mathbf{A}$ são estimados por (2.17).

Portanto ao estimar os pesos ótimos a variância obtida é mínima. Deste resultado segue que ao se comparar a variância de um estimador com p variáveis auxiliares com outro com $p + 1$ variáveis, a variância do estimador com mais variáveis será menor.

Uma aplicação do Estimador do Tipo Razão Multivariado com duas variáveis auxiliares é obtido através do exemplo utilizado por Olkin (1958). Deseja-se estimar o número de habitantes de 200 cidades grandes do Estados Unidos em 1930, excluindo as cinco maiores. Onde $Y = 1.950$, $X_1 = 1.940$ e $X_2 = 1.930$. Uma amostra de tamanho 50 é obtida e $\overline{Y} = 1.699$ e $\sigma(\overline{Y}) = 1.740$, $\overline{X}_1 = 1.482$ e $\sigma(\overline{X}_1) = 1.554$, $\overline{X}_2 = 1.420$ e $\sigma(\overline{X}2) = 1.509$, $\overline{y} = 1.896$ e $s(\overline{y}) = 2.088$, $\overline{x}_1 = 1.693$ e $s(\overline{x}_1) = 1.932$, $\overline{x}_2 = 1.643$ e $s(\overline{x}_2) = 1.931$.

$$\mathbf{e_{pop}} = \begin{pmatrix} 1,049 & 1,059 & 1,056 \\ & 1,098 & 1,108 \\ & & 1131 \end{pmatrix}$$

$$\mathbf{e_{am}} = \begin{pmatrix} 1,213 & 1,241 & 1,256 \\ & 1,302 & 1,335 \\ & & 1,381 \end{pmatrix}$$

$$\mathbf{A_{pop}} = \begin{pmatrix} 0,029 & 0,042 \\ & 0,068 \end{pmatrix}$$

$$\mathbf{A_{am}} = \begin{pmatrix} 0,033 & 0,051 \\ & 0,082 \end{pmatrix}$$

$$\mathbf{b_{pop}} = \begin{pmatrix} 0,039 & 0,075 \end{pmatrix}$$

$$\mathbf{b_{am}} = \begin{pmatrix} 0,061 & 0,125 \end{pmatrix}$$

$$\hat{\mathbf{w}}_{\mathbf{pop}} = \begin{pmatrix} 2 & -1 \end{pmatrix}$$

$$\hat{\mathbf{w}}_{\mathbf{am}} = \begin{pmatrix} 2,38 & -1,38 \end{pmatrix}$$

O Estimador do Tipo Razão Univariado de $\overline{Y}$ é dado por:

$$\hat{\overline{Y}} = r\overline{X} = \frac{1.896}{1.693}1.482 = 1.660 \tag{2.23}$$

$$\sigma(\hat{\overline{Y}}) = 289f \tag{2.24}$$

onde $f^2 = \dfrac{(N-n)}{nN}$

O Estimador do Tipo Razão Bivariado de $\overline{Y}$ é dado por:

(1)Com pesos reais

$$\hat{\overline{Y}} = 2r_1\overline{X}_1 - r_2\overline{X}_2 = 1.681 \tag{2.25}$$

$$\sigma(\hat{\overline{Y}}) = 2,7f \tag{2.26}$$

(2)Com pesos estimados

$$\hat{\overline{Y}} = 2,38r_1\overline{X}_1 - 1,38r_2\overline{X}_2 = 1.689 \tag{2.27}$$

Uma vez que c_x tem valor próximo de c_y e $\rho_{x_1y} = 0,987$, o Estimador do Tipo Razão tem valor maior que a média estimada e o estimador com duas variáveis auxiliares é melhor que o estimador com um variável auxiliar.

Capítulo 3

ESTIMADOR DO TIPO REGRESSÃO

3.1 ESTIMADOR DO TIPO REGRESSÃO UNIVARIADO

Como visto no Capítulo anterior, para X e Y obedecendo a uma relação linear passando pela origem, Estimadores do Tipo Razão seriam mais adequados do que os estimadores simples. Por outro lado, estudando-se a relação X e Y, pode-se concluir que, embora linear, ela não passa pela origem. Isto sugere um estimador baseado na regressão de Y em X, e não na razão de duas variáveis. A variável auxiliar X é supostamente conhecida pela população, bem como no Estimador do Tipo Razão. Portanto o Estimador do Tipo Razão é um caso particular do Estimador do Tipo Regressão que tem a média da população dada por (Hanuschak, 1978):

$$\hat{\overline{Y}} = \overline{y} + \hat{b}(\overline{X} - \overline{x}) \tag{3.1}$$

onde $\hat{b}$ é uma estimativa do impacto em Y provocado pela variação de uma unidade na variável X, sendo estimado por:

$$\hat{b} = \frac{\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})}{\sum_{i-1}^{n}(x_i - \overline{x})^2} = \frac{sxy}{s_x^2} \tag{3.2}$$

A variância de $\hat{\overline{Y}}$ é dado por:

$$Var(\hat{\overline{Y}}) = fS_Y^2(1 - \rho^2) \tag{3.3}$$

no qual $f = \dfrac{N-n}{nN}$, e ρ o coeficiente de correlação entre X e Y.

3.2 ESTIMADOR DO TIPO REGRESSÃO MULTIVARIADO

Hanuschak (1978) considerou p variáveis auxiliares independentes para a estimação de Y. O Estimador do Tipo Regressão Multivariado para a média populacional será dado por:

$$\hat{\overline{Y}} = \overline{y} + \sum_{i=1}^{p} B_i(\overline{X}_i - \overline{x}_i) \tag{3.4}$$

e na forma matricial:

$$\hat{\overline{Y}} = \overline{y} + \mathbf{B}'(\overline{\mathbf{X}} - \overline{\mathbf{x}}) \tag{3.5}$$

O estimador $\hat{\overline{Y}}$ é não viesado se $\mathbf{B}'$ é fixo e possui variância dada por:

$$Var(\hat{\overline{Y}}) = f(S_y^2 + \mathbf{B}'\mathbf{\Sigma_{XX}}\mathbf{B} - \mathbf{2B}'\mathbf{\Sigma_{XY}}) \tag{3.6}$$

$$\mathbf{\Sigma_{XX}} = \begin{pmatrix} S_{X_1}^2 & S_{X_1X_2} & \dots & S_{X_1X_P} \\ S_{X_2X_1} & S_{X_2}^2 & \dots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ S_{X_PX_1} & S_{X_PX_2} & \dots & S_{X_2}^2 \end{pmatrix}$$

$$\mathbf{\Sigma_{XY}} = \begin{pmatrix} S_{X_1Y} \\ S_{X_2Y} \\ \vdots \\ S_{X_PY} \end{pmatrix}$$

onde $f = \dfrac{N-n}{nN}$, Σ_{XX} é a matriz de variância e covariância das variáveis auxiliares e Σ_{XY} a matriz de correlação entre variável de interesse Y e às variáveis auxiliares X_i's. O vetor $\mathbf{B}$ é obtido de tal forma que a variância de $\hat{\overline{Y}}$ seja minimizada, então a equação é derivada com relação a $\mathbf{B}$ e em seguida igualada a zero. Obtem-se o seguinte resultado:

$$\hat{\mathbf{B}} = \mathbf{\Sigma_{XX}^{-1}}\mathbf{\Sigma_{XY}} \tag{3.7}$$

Portanto a variância minimizada de $\hat{\overline{Y}}$ é dada por:

$$Var(\hat{\overline{Y}}) = fS_Y^2(1 - R^2_{YX_1X_2\dots X_P}) \tag{3.8}$$

onde $R^2_{YX_1X_2...X_P}$ é o coeficiente de correlação múltipla populacional definida por Anderson (1958):

$$R^2_{YX_1X_2...X_P} = \frac{\sqrt{\hat{\mathbf{B}}'\mathbf{\Sigma_{XX}}\hat{\mathbf{B}}}}{S_Y} \tag{3.9}$$

Observe que a variância de $\hat{\bar{Y}}$ para o caso multivariado e o caso univariado se diferem apenas pelo tipo de coeficiente de correlação.

Capítulo 4

MATERIAIS E MÉTODOS

Para o estudo serão criados 5 conjuntos de dados ultizando o *software* SAS. Cada conjunto é composto pela variável de interesse Y, pelas variáveis auxiliares $X_i's$ e pelos $b_i's$ que é o produto entre o peso w_i e a razão (R_i) entre a média da variável de interesse $\overline{Y}$ e a média da variável auxiliar $\overline{X}_i$. Desses conjuntos de dados serão utilizados uma quantidade diferente de variáveis auxiliares para a determinação de Y da seguinte forma:

```
data ax1;
do j=1 to 100;
do i=1 to 100;
 x1=rannor(i*j)+2;
 b1=rannor(sqrt(i*j))*0.006+0.6;
 y=b1*x1;
output;
end;
end;
run;

data ax12;
do j=1 to 100;
do i=1 to 100;
 x1=rannor(i*j)+2;
 x2=rannor(i*j)+5;
 b1=rannor(sqrt(i*j))*0.007+0.7;
 b2=rannor(sqrt(i*j))*0.003+0.3;
 y=b1*x1+b2*x2;
output;
end;
end;
run;
```

```
data ax123;
do j=1 to 100;
do i=1 to 100;
 x1=rannor(i*j)+2;
 x2=rannor(i*j)+5;
 x3=rannor(i*j)+3;
 b1=rannor(sqrt(i*j))*0.005+0.5;
 b2=rannor(sqrt(i*j))*0.002+0.2;
 b3=rannor(sqrt(i*j))*0.003+0.3;
 y=b1*x1+b2*x2+b3*x3;
output;
end;
end;
run;

data ax1234;
do j=1 to 100;
do i=1 to 100;
 x1=rannor(i*j)*1+2;
 x2=rannor(i*j)*2+5;
 x3=rannor(i*j)+3;
 x4=rannor(i*j)*2+4;
 b1=rannor(sqrt(i*j))*0.005+0.5;
 b2=rannor(sqrt(i*j))*0.002+0.2;
 b3=rannor(sqrt(i*j))*0.001+0.1;
 b4=rannor(sqrt(i*j))*0.002+0.2;
 y=b1*x1+b2*x2+b3*x3+b4*x4;
output;
end;
end;
run;

data ax12345;
do j=1 to 100;
do i=1 to 100;
 x1=rannor(i*j)*1+2;
 x2=rannor(i*j)*2+5;
 x3=rannor(i*j)+3;
 x4=rannor(i*j)*2+4;
 x5=rannor(i*j)*6+10;
 b1=rannor(sqrt(i*j))*0.005+0.5;
 b2=rannor(sqrt(i*j))*0.002+0.2;
 b3=rannor(sqrt(i*j))*0.001+0.1;
 b4=rannor(sqrt(i*j))*0.0013+0.13;
 b5=rannor(sqrt(i*j))*0.0007+0.07;
 y=b1*x1+b2*x2+b3*x3+b4*x4+b5*x5;
output;
end;
end;
run;
```

Ou seja, serão criados 100 conjunto de dados com tamanho 100 para cada base de dados. A idéia da criação de 5 conjunto de dados se deve ao fato de que com 4 variáveis auxiliares na determinação de Y, por exemplo, terá-se o Estimador do

Tipo Razão. Se deste conjunto de dados retirar-se 1 ou mais variáveis, o Estimador do Tipo Regressão será mais adequado, pois o Estimador do Tipo Razão é um caso particular do Regressão, com o intercepto igual a zero. Portanto, neste caso, ao se retirar 1 variável auxiliar teremos uma combinação 4 a 3 de Estimadores do Tipo Regressão, se retirar 2, uma combinação 4 a 2 e assim por diante. Através dessas combinações, irá ser feita uma comparação com o Estimador do Tipo Razão, comparando a média e a variância de Y para saber qual estimador é mais eficiente. O mesmo processo será repetido para os outros conjuntos de dados.

Portanto primeiramente será feita uma comparação entre Estimador do Tipo Razão Multivariado com todas as p variáveis auxiliares da base de dados e o Estimador do Tipo Regressão com no máximos $p-1$ variáveis auxiliares para saber qual estimador é mais eficiente. Será utilizado a seguinte macro para gerar os estimadores:

```
%macro mr(data=,y=,x=,xreg=,My=,Mx=,Mxreg=,n=);
proc iml;
print "X: &x";
use &data;
read all var {&y} into y;
read all var {&x} into x;
read all var {&xreg} into xr;
p=ncol(x);
n=nrow(x);
My=&My;
Mx={&Mx};
Mxr={&Mxreg};
print My Mx Mxr;

A=j(p,p,0);
do i=1 to p;
do j=1 to p;
ri=y[:]/x[:,i];
rj=y[:]/x[:,j];
A[i,j]=(y-ri*x[,i])`*(y-rj*x[,j])/(n-1);
end;
end;
print A;

r=j(1,p,0);
do i=1 to p;
r[i]=y[:]/x[:,i];
end;

xb=j(1,ncol(xr),0);
do i=1 to ncol(xr);
xb[i]=xr[:,i];
end;

detA=det(A);*print detA;
e=j(1,p,1);
```

```
w=e*inv(A)*(inv(e*inv(A)*e‘));
sw=w[+];
print w[colname={&x}] sw;
print r[colname={&x}];
b=(w#r);
print b[colname={&x}];
Yr=(w#r)*Mx‘;
yb=y[:];
xr1=j(nrow(xr),1,1)||xr;
beta=inv(xr1‘*xr1)*xr1‘*y;print beta;
yreg=yb+beta[2:ncol(xr)+1]‘*(Mxr-xb)‘;
print Yr Yreg yb My;
nn=&n;
VarYr=(1-n/100)/(n*(e*inv(A)*e‘));
VarAAS=(1-n/100)*(y-y[:])‘*(y-y[:])/((n-1)*n);
print VarYr VarAAS;
create estimates var{Yr Yreg yb My VarYr VarAAS};append;
quit;
proc surveymeans data=&data var total=100;
var &y;
run;
%mend mr;
```

Esta é a PROC IML do Estimator do Tipo Razão implementada do software SAS. Também é possível fazer a comparação entre o Estimador do Tipo Razão com p variáveis e o Estimador do Tipo Regressão com p variáveis, para isto basta fazer a entrada das variáveis auxiliares X_i do Estimador Razão como Estimador Regressão.

Então com esses cálculos será possível estimar os pesos $w_i's$, a média populacional, amostral e do estimador razão, bem como a variância amostral e do estimador razão, os $\beta_i's$, as razões $R_i's$ e a matriz $\mathbf{A}$. Para obter as estimativas será utilizado o PROC SURVEYSELECT no SAS. Segue exemplo de um programa para uma base de dados com 4 variáveis auxiliares e uma amostra de tamanho 30:

```
proc surveyselect data=ax1234(where=(j=1)) out=ax12341 sampsize=30
seed=2 noprint;run;
proc sql noprint;
select mean(y) into:My from ax1234 where j=1;
select mean(x1) into:Mx1 from ax1234 where j=1;
select mean(x2) into:Mx2 from ax1234 where j=1;
select mean(x3) into:Mx3 from ax1234 where j=1;
select mean(x4) into:Mx4 from ax1234 where j=1;
quit;%put &my &mx1 &mx2 &mx3 &mx4 R1=%sysevalf(&my/&mx1)
R2=%sysevalf(&my/&mx2) R3=%sysevalf(&my/&mx3) R4=%sysevalf(&my/&mx4);
%mr(data=ax12341,y=y,x=x1 x2 x3 x4,My=&my,Mx=&mx1 &mx2 &mx3 &mx4,n=30);
```

Após a obtenção das estimativas pontuais do Estimador do Tipo Razão Multivariado é necessário a sua comparação com o Estimador do Tipo Regressão Multivariado. Serão utilizados vários tamanhos de amostras, iniciado com tamanho 10, depois com incrementos de 10 unidades amostrais até possuir uma amostra do tamanho de uma população, ou seja, 100 unidades amostrais. Serão criados então 100 amostras para cada conjunto de dados e cada tamanho de amostra. A comparação será através do menor viés e da menor variância. Segue um exemplo de uma comparação com todas as combinações possíveis de variáveis auxiliares para a quarta base de dados:

```
%simula(tab=ax1234,x=x1 x2 x3 x4,xreg=x1,nvar=x1);
%simula(tab=ax1234,x=x1 x2 x3 x4,xreg=x2,nvar=x2);
%simula(tab=ax1234,x=x1 x2 x3 x4,xreg=x3,nvar=x3);
%simula(tab=ax1234,x=x1 x2 x3 x4,xreg=x4,nvar=x4);
%simula(tab=ax1234,x=x1 x2 x3 x4,xreg=x1 x2,nvar=x12);
%simula(tab=ax1234,x=x1 x2 x3 x4,xreg=x1 x3,nvar=x13);
%simula(tab=ax1234,x=x1 x2 x3 x4,xreg=x1 x4,nvar=x14);
%simula(tab=ax1234,x=x1 x2 x3 x4,xreg=x2 x3,nvar=x23);
%simula(tab=ax1234,x=x1 x2 x3 x4,xreg=x2 x4,nvar=x24);
%simula(tab=ax1234,x=x1 x2 x3 x4,xreg=x3 x4,nvar=x34);
%simula(tab=ax1234,x=x1 x2 x3 x4,xreg=x1 x2 x3,nvar=x123);
%simula(tab=ax1234,x=x1 x2 x3 x4,xreg=x1 x2 x4,nvar=x124);
%simula(tab=ax1234,x=x1 x2 x3 x4,xreg=x1 x3 x4,nvar=x134);
%simula(tab=ax1234,x=x1 x2 x3 x4,xreg=x2 x3 x4,nvar=x234);
%simula(tab=ax1234,x=x1 x2 x3 x4,xreg=x1 x2 x3 x4,nvar=x1234);
```

Este programa irá fornecer tabelas com as estimativas e gráficos com comparações. Será feito ainda uma contagem para saber em quantas amostras qual Estimador foi mais eficiente. Para título de comparação será utilizado também o Estimador da Amostra Aleatória Simples (AAS).

Capítulo 5

ANÁLISE DOS RESULTADOS

A análise dos resultados será dividida em duas partes. Na primeira será feita uma verificação do algoritmo proposto por Olkin (1958), obtendo as estimativas pontuais. Na segunda parte será realizada a comparação entre os Estimadores do Tipo Razão, do Tipo Regressão e AAS.

5.1 ESTIMATIVAS PONTUAIS

Primeiramente será exposto as estimativas pontuais através do algoritmo exibido anteriormente. Os resultados para o conjunto de dados com 4 variáveis auxiliares de uma amostra de tamanho 30 são:

$$\mathbf{A} = \begin{pmatrix} 1.4657 & -1,1249 & -0,2031 & -0,4815 \\ & 1,6902 & -0,0528 & -0,4826 \\ & & 1,3796 & -0,1786 \\ & & & 1,2041 \end{pmatrix}$$

$$\mathbf{w} = \begin{pmatrix} 0,3335 & 0,3019 & 0,0957 & 0,2688 \end{pmatrix}$$

Então foi obtido as matrizez $\mathbf{A}$ e $\mathbf{w}$. Note que a soma do pesos é igual a 1, como foi prosposto na teoria, de forma que produziu a variância mínima para o conjunto de dados utilizando o Estimador do Tipo Razão. O maior peso foi atribuído à variável X_1 apesar de possuir a menor média(2), porém é a que mais contribui para estimação de Y com $b_1 = 0,5$. Já a variável de menor peso é X_3 que possui a segunda menor média (3) e que menos contribui $b_3 = 0,1$. A variável X_2 é a que possui maior média(5), porém não é a que contribui mais, pois $b_2 = 0,2$, mesmo assim tem peso considerável. Já a variável X_4 tem a segunda maior média (4) e contribui o mesmo que X_2, mas com peso um pouco menor.

$$\mathbf{R} = \begin{pmatrix} 1,5088 & 0,6673 & 1,0208 & 0,7383 \end{pmatrix}$$

Como era de se esperar, quanto maior a média da variável X_i menor será a razão R_i.

$$\mathbf{b} = \begin{pmatrix} 0,5032 & 0,2015 & 0,0977 & 0,1985 \end{pmatrix}$$

A matriz **b** foi obtida através do produto das matrizes **w** e **R**. Como pode ser observado, foi obtido os valores definidos anteriormente o que implica que o algoritmo foi desenvolvido corretamente.

$\overline{Y}_{RAZ}$	$\overline{Y}_{AAS}$	$\overline{Y}_{POP}$
3,0352	3,2047	3,0363

VAR_{raz}	VAR_{AAS}
0,000007891	0,0101708

Pode ser observado que para um tamanho de amostra 30, o Estimador do Tipo Razão da média de Y foi mais preciso do que o estimador amostral (AAS), pois a média obtida está mais próxima da média populacional, e a variância é bem menor que da Amostra Aleatória Simples (AAS) com $DEFF = 0,000776$. Em resumo, o Estimador do Tipo Razão para este caso é muito mais adequado para o plano amostral do que o Estimador AAS.

5.2 COMPARAÇÃO DOS ESTIMADORES DO TIPO RAZÃO E REGRESSÃO MULTIVARIADOS

Nesta seção serão apresentados os resultados das comparações entre os Estimadores do Tipo Razão, do Tipo Regressão e do AAS. Será exposto os resultados da quinta base de dados. Não será necessário a apresentação dos resultados das outras bases de dados porque os mesmos são bem semelhantes ao do conjunto de dados que será exposto. Em um primeiro momento o Estimador do Tipo Razão terá as $p = 5$ variáveis auxiliares na estimação e o Estimador do Tipo Regressão todas as combinações possíveis com no máximo $p - 1$ variáveis auxiliares. A idéia é verificar se o Estimador Razão é mais eficiente com pelo menos uma variável auxiliar a mais que o Regressão. No segundo momento a comparação será realizada com a mesma quantidade de variáveis auxiliares tanto para o Razão quanto para o Regressão.

5.2.1 P VARIÁVEIS PARA O ESTIMADOR DO TIPO RAZÃO E $(P - 1)$ VARIÁVEIS PARA O ESTIMADOR DO TIPO REGRESSÃO

Foi utilizado 100 amostras para cada tamanho de amostra. Ela foi iniciada com tamanho 10 e outras com incrementos de 10 unidades amostrais, até a amostra ter o tamanho da população (100). Então foram obtidas 1.000 amostras. Foi feito gráficos e tabelas da estimativas da média de y e da variância de $\overline{Y}$.

Para a variância, o Estimador do Tipo Razão foi o melhor em todas as amostras, exceto quando a amostra era do tamanho da população em que a variância de todos os Estimadores são iguais a zero. Segue o gráfico 5.1 no qual ilustra o comportamento da variância dos 3 estimadores para uma Regressão com as variáveis auxiliares X_1, X_3 e X_4:

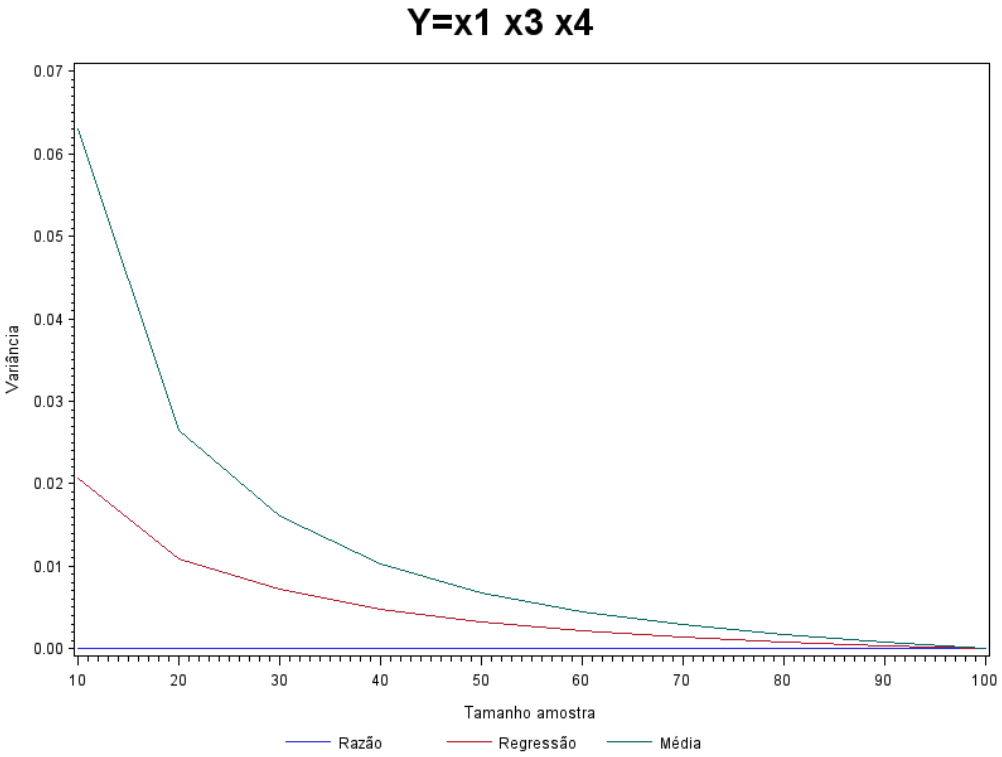

Figura 5.1: VAR - Regressão $X_1X_3X_4$

Perceba que a variância do Estimador AAS é a maior e que a do Estimador do Tipo Regressão, que está entre a ASS e a do Razão. A variância do Estimador do Tipo Razão está bem próxima de zero.

Para o viés de $\overline{Y}$ o Estimador do Tipo Razão foi melhor em mais de 95% das amostras. A Tabela 5.1 apresenta a frequência de vezes em que cada Estimador foi o melhor em cada amostra:

Tabela 5.1: Frequência para o menor Viés

Regressão	AAS	Razão	Regressão	Total
x1	18	861	21	900
x2	17	868	15	900
x3	15	865	20	900
x4	18	859	23	900
x5	18	859	23	900
x1x2	17	871	12	900
x1x3	18	868	14	900
x1x4	18	862	20	900
x1x5	17	863	20	900
x2x3	17	867	16	900
x2x4	18	867	15	900
x2x5	18	868	14	900
x3x4	17	869	14	900
x3x5	17	865	18	900
x4x5	18	861	21	900
x1x2x3	18	857	25	900
x1x2x4	18	858	24	900
x1x2x5	18	853	29	900
x1x3x4	18	858	24	900
x1x3x5	18	854	28	900
x1x4x5	18	854	28	900
x2x3x4	18	866	16	900
x2x3x5	18	859	23	900
x2x4x5	17	870	13	900
x3x4x5	18	867	15	900
x1x2x3x4	16	860	24	900
x1x2x3x5	18	841	41	900
x1x2x4x5	15	788	97	900
x1x3x4x5	18	849	33	900
x2x3x4x5	17	867	16	900
Total	524	25774	702	27000

Contudo para cada regressão foram contabilizadas 900 amostras e não 1000 como foi dito. Isto se deve ao fato de que nas 100 amostras do tamanho da população (100), não têm viés. Portanto todos os estimadores são iguais a média populacional. A Figura 5.2 apresenta o comportamento do viés dos estimadores para a média de Y:

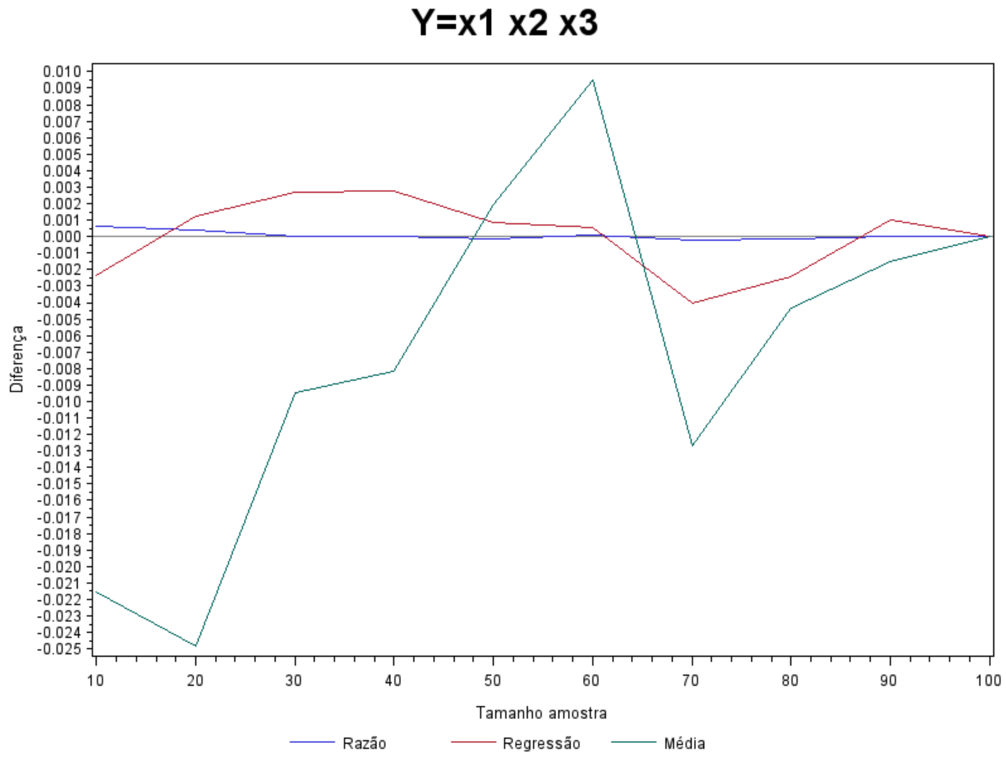

Figura 5.2: VIÉS - Regressão $X_1X_2X_3$

Observe que quando a amostra é do tamanho da população, os estimadores deixam de ter viés. Veja ainda que o Estimador do Tipo Razão é o que tem menor viés. Já o Regressão, neste caso, tem viés menor que do AAS.

Portanto o Estimador do Tipo Razão é mais eficiente quando ele tem todas as variáveis auxiliares na sua estimação, comparado ao Estimador do Tipo Regressão com 1 variável a menos no máximo.

5.2.2 ESTIMADORES DO TIPO RAZÃO E REGRESSÃO COM A MESMA QUANTIDADE DE VARIÁVEIS AUXILIARES

Assim como na subseção anterior, o mesmo procedimento para a criação de amostra foi adotado, porém as estimativas para o Estimado do Tipo Razão e do Tipo Regressão possuem as mesma variáveis auxiliares na estimação. Desta forma a variância de $\overline{Y}$ foi menor para o Estimador do Tipo Regressão em todos os modelos. O gráfico 5.3 mostra o comportamento da variância para o modelo igual a X_4X_5:

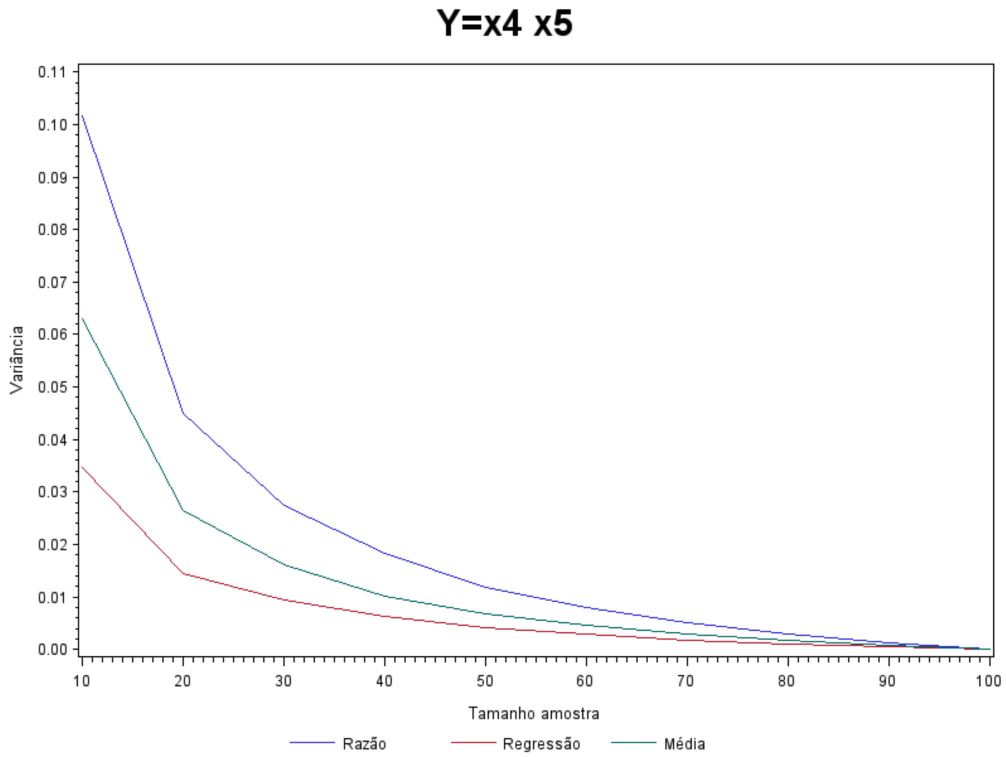

Figura 5.3: VAR - Modelo X_4X_5

Já o gráfico 5.4 mostra o comportamento da variância para o modelo igual a $X_1X_2X_3$:

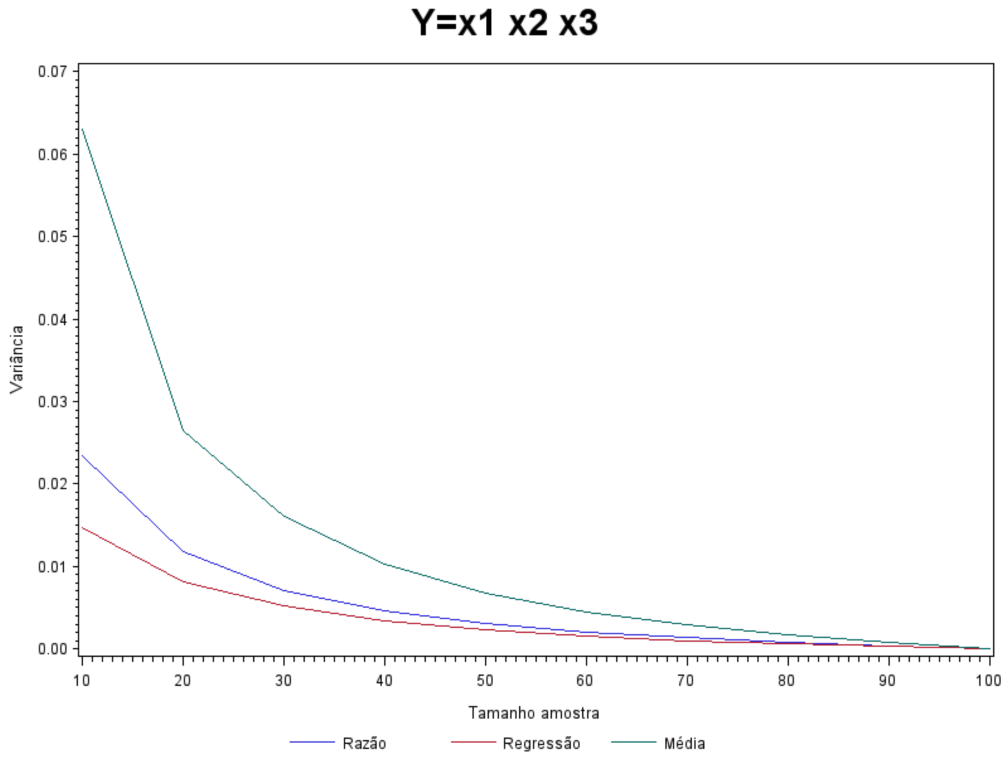

Figura 5.4: VAR - Modelo $X_1X_2X_3$

Como pode ser observado na Figura 5.3 a variância do Estimado do Tipo Razão é pior que o Estimador AAS, já na Figura 5.4 é o contrário. Para entender melhor este comportamento, a Tabela 5.2 mostra a frequência de vezes em que cada Estimador foi melhor em todas as amostras criadas.

Tabela 5.2: Frequência para a menor Variância - AAS x Razão

Modelo	AAS	Razão	Total
x1	890	10	900
x2	878	22	900
x3	891	9	900
x4	897	3	900
x5	897	3	900
x1x2	111	789	900
x1x3	588	312	900
x1x4	669	231	900
x1x5	538	362	900
x2x3	645	255	900
x2x4	650	250	900
x2x5	455	445	900
x3x4	834	66	900
x3x5	798	102	900
x4x5	846	54	900
x1x2x3	17	883	900
x1x2x4	8	892	900
x1x2x5	0	900	900
x1x3x4	128	772	900
x1x3x5	35	865	900
x1x4x5	18	882	900
x2x3x4	212	688	900
x2x3x5	81	819	900
x2x4x5	45	855	900
x3x4x5	372	528	900
x1x2x3x4	1	899	900
x1x2x3x5	0	900	900
x1x2x4x5	0	900	900
x1x3x4x5	3	897	900
x2x3x4x5	9	891	900
x1x2x3x4x5	0	900	900
Total	11516	16384	27900

Como pode ser visto na Tabela 5.2, a variância do Estimador do Tipo Razão foi menor que o AAS em 59% das amostras. Veja que quando o Razão tinha apenas uma variável auxiliar, o Estimador AAS ganhou em quase todas as amostra. Já o Razão passou a ser superior a medida que outra variável fosse adicionada ao modelo. Este fato se deve porque ao se adicionar uma variável, o modelo passa a ter mais informação, se tornando assim melhor que o AAS.

Agora será apresentado o resultado do viés de $\overline{Y}$ para os três estimadores. A tabela 5.3 mostra a frequência de vezes em que cada Estimador foi melhor em cada amostra.

Tabela 5.3: Frequência para o menor Viés - AAS x Razão x Regressão

Modelo	AAS	Razão	Regressão	Total
x1	342	224	334	900
x2	364	271	265	900
x3	398	257	245	900
x4	402	222	276	900
x5	402	222	276	900
x1x2	281	321	298	900
x1x3	310	269	321	900
x1x4	316	289	295	900
x1x5	276	269	355	900
x2x3	326	310	264	900
x2x4	344	318	238	900
x2x5	305	322	273	900
x3x4	370	291	239	900
x3x5	318	269	313	900
x4x5	309	281	310	900
x1x2x3	280	330	290	900
x1x2x4	249	352	299	900
x1x2x5	288	316	396	900
x1x3x4	307	282	311	900
x1x3x5	264	281	355	900
x1x4x5	239	310	351	900
x2x3x4	345	332	223	900
x2x3x5	287	308	305	900
x2x4x5	293	343	264	900
x3x4x5	308	299	293	900
x1x2x3x4	253	341	306	900
x1x2x3x5	166	353	381	900
x1x2x4x5	58	328	514	900
x1x3x4x5	227	316	357	900
x2x3x4x5	281	350	269	900
x1x2x3x4x5	18	471	411	900
Total	8826	9447	9627	27900

Como pode ser observado na Tabela 5.3, há um certo equilíbrio sobre o viés com uma pequena vantagem para o Estimador do Tipo Regressão, que foi melhor em 35% das amostras. Veja ainda que ao se utilizar 1 variável auxiliar, o Estimador

AAS foi melhor que o do Tipo Razão e do Tipo Regressão. Ao se adicionar variáveis auxiliares no modelo a situação se inverte. Isto acontece porque a medida que se insere uma variável, mais informação estará continda no modelo, tornando assim o Estimadores do Tipo Razão e do Tipo Regressão melhores.

O certo equilíbrio mencionado pode ser visto claramente na Figura 5.5.

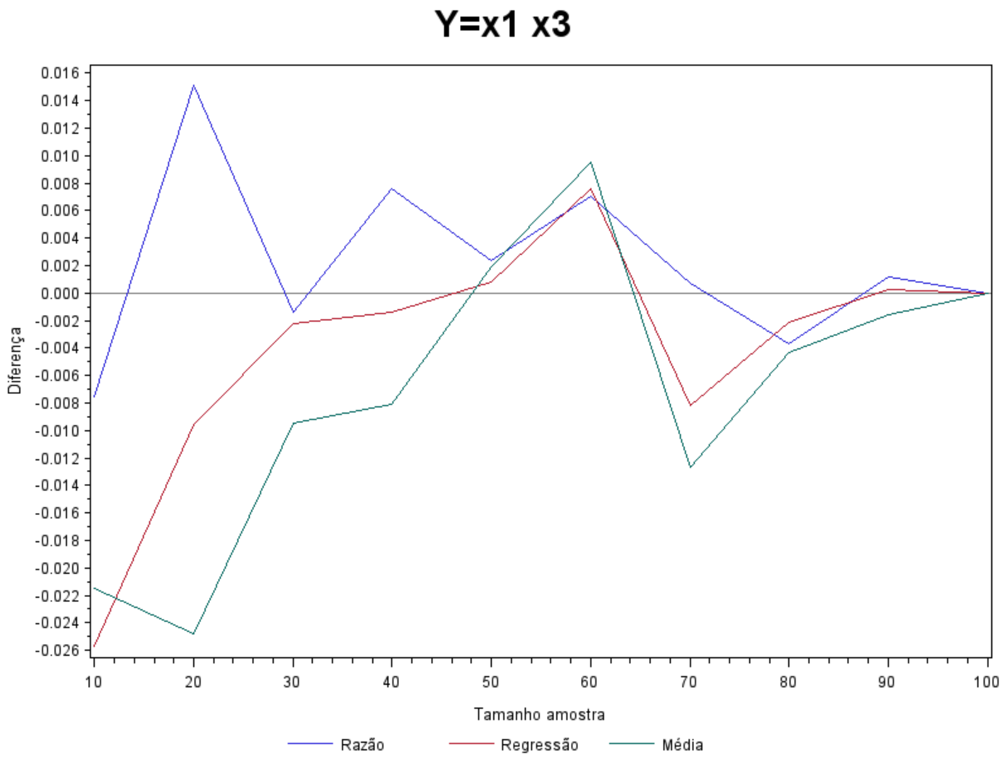

Figura 5.5: VIÉS - Modelo X_1X_3

Observando a Figura 5.5 o viés é maior para pequenas amostras, contudo ele se estabiliza para amostras a partir de tamanho 30. Segue agora a tabela 5.4 que mostra a frequências de vezes em que o Estimador do Tipo Razão e do Tipo Regressão foram melhores.

Tabela 5.4: Frequência para o menor Viés - Razão x Regressão

Modelo	Razão	Regressão	Total
x1	224	676	900
x2	271	629	900
x3	258	642	900
x4	223	667	900
x5	223	677	900
x1x2	346	554	900
x1x3	326	574	900
x1x4	316	584	900
x1x5	287	613	900
x2x3	345	555	900
x2x4	345	555	900
x2x5	346	554	900
x3x4	324	576	900
x3x5	310	590	900
x4x5	304	596	900
x1x2x3	397	503	900
x1x2x4	386	514	900
x1x2x5	350	550	900
x1x3x4	351	549	900
x1x3x5	337	563	900
x1x4x5	348	552	900
x2x3x4	395	505	900
x2x3x5	359	541	900
x2x4x5	383	517	900
x3x4x5	351	549	900
x1x2x3x4	418	482	900
x1x2x3x5	401	499	900
x1x2x4x5	346	554	900
x1x3x4x5	379	521	900
x2x3x4x5	407	493	900
x1x2x3x4x5	480	420	900
Total	10536	17364	27900

Como pode ser observado na Tabela 5.4, o viés do Estimador do Tipo Regressão é sempre menor que o do Tipo Razão em 62% das amostras quando as mesmas variáveis estão no modelo. O que pode ser claramente observado é que a medida que se acrescenta uma variável auxiliar no modelo, essa diferença diminui. Pode ser observado ainda que ao se utilizar todas as variáveis no modelo, o víés do Estimador do Tipo Razão é menor que o do Tipo Regressão em 53% das amostras, mostrando que quando se tem todas as variáveis disponíveis, o Estimador Razão possa ser mais adequado. Isso pode ser corrobrado ao ver o gráfico 5.6, que mostra o comportamento da variância quando todas as variáveis estão no modelo, pois a diferença é mínima. Esse resultado também é obtido ao utilizar o Estimador do Tipo Regressão com a restrição do intercepto igual a zero.

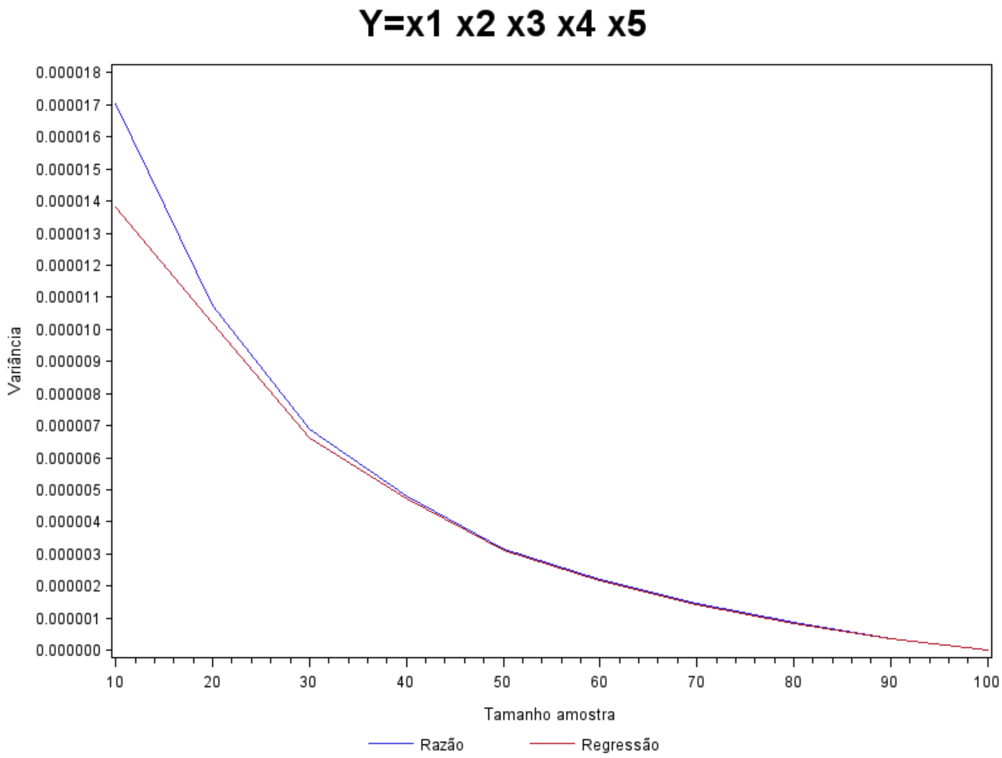

Figura 5.6: VAR - Modelo $X_1X_2X_3X_4X_5$

Portanto o Estimador do Tipo Regressão é uma escolha melhor que o do Tipo Razão quando as mesmas variáveis auxiliares são utilizadas no modelo. Contudo ao se utilizar todas as variáveis auxiliares no modelo o Estimador do Tipo Razão pode ser uma melhor escolha. Outro ponto que pode ser destacado é que quando os modelos possuem a mesma quantidade de variáveis auxiliares, a variância do Regressão é sempre menor que a variância da Razão, assim como no caso univariado.

Capítulo 6

CONCLUSÕES

O estudo em questão foi motivado por haver pouca discussão sobre o assunto e por não haver a implementação do Estimador do Tipo Razão Multivariado no SAS. Portanto com este trabalho foi possível verificar as suposições quanto ao estimador proposto por Olkin (1958) em seu artigo.

Após a confirmação do algoritmo é necessário o estudo de sua eficiência. Nada mais conveniente do que comparar com o Estimador do Tipo Regressão Multivariado, que já tem sua implementação no SAS via PROC REG. Esta comparação é importante, pois assim como no caso univariado o Estimador Razão é um caso particular do Estimador Regressão. O Estimador AAS também foi levado em conta devido a sua simplicidade na estimação. Caso este estimador fosse pelo menos tão eficiente quanto o Razão ou o Regressão, seria mais viável a sua utilização. Porém somente em casos excepcionais foi observado que o AAS foi melhor, pois no geral os Estimadores do Tipo Razão e do Tipo Regressão foram mais eficientes.

Então no estudo ficou claro que tanto o Estimador do Tipo Razão ou o Estimador do Tipo Regressão pode ser mais eficiente. O Razão será sempre melhor quando possuir um número maior de variáveis auxiliares que o Regressão. Este último será melhor quando possuir a mesma quantidade de variáveis que o Razão, exceto no caso em que todas as variáveis auxiliares estão no modelo, podendo o Razão também ser adequado. Portanto a utilização de quaisquer estimadores estudados irá depende apenas do delineamento do problema.

Referências Bibliográficas

Anderson, T. W. (1958). An introduction to multivariate statistical analysis. *Jonh Wiley and Sons.*

Bolfarine, H. & Bussab, W. O. (2005). *Elementos de Amostragem.* ABE - Projeto Fisher.

Cochran, W. G. (1977). *Sampling Techniques*, (3rd ed.). John Wiley & Sons.

Hanif, M., Ahmad, Z., & Ahmad, M. (2009). Generalized multivariate ratio estimators using multi-auxiliary variables for multi-phase sampling. *Great Socialist Peoples Libyan Arab Jamahiriya.* p.647 - 661.

Hanuschak, G. A. (1978). Multiple regression estimation using classifield landsat data. Technical report, U. S. Department of Agriculture.

Hartley, H. O. & Ross, A. (1954). Unbiased ratio estimators. *Nature, London, 174, no. 4423, 270.*

Lira, S. A. (2004). Análise de correlação: Abordagem teórica e de construção dos coeficientes com aplicações. Master's thesis, Universidade Federal do Paraná.

Ngesa, O. O., Orwa, G. O., Otieno, R. O., & Murray, H. M. (2012). Multivariate ratio estimator of the population total under stratifield random sampling. *Journal of Statistics.* p. 300 - 304.

Olkin, I. (1958). Multivariate ratio estimation for finite populations. *Biometrika Trust.* p. 154 - 165.

Singh, H. P., Kumar, S., & Bhougal, S. (2011). Multivariate ratio estimation in presence of non-response in successive sampling. *Journal os Statistical Theory and Practice.* p. 591 - 611.

Singh, R., Malik, S., Adewara, A. A., & Smarandache, F. (2012). Multivariate ratio estimation with known population proportion of two auxiliary characters for finite population. *viXra:1204.0077.*